Mathematical Principles of Machine Learning, Spring 2019

Overview and Logistics

I-Hsiang Wang
National Taiwan University
ihwang@ntu.edu.tw

2019.02.19

From Data to Intelligence

data → **Machine Learning** → **AI**

"AI" = "skills"
in most case

- Many successful applications needless to say …
- How well a machine can learn depends on many factors.
  - What learning model should be chosen?
  - How to train the learning model?
- What to do when the performance sucks?
  - insufficient amount of data? bad quality of data?
  - inappropriate ML model? bad training algorithm? wrong features?
- Theory behind the design of ML?

What to Expect from this Course

- Common wisdom in ML:
  Good test performance
  = Good training performance + Good generalization

- Training performance: algorithmic principle
- Generalization: statistical principle

- Goal of this course: understand the foundations of machine learning with solid theoretical development

- Make a wish: what do you want a good theory to tell you?
- At the end of the semester, check if the current theory meets your wish.

Objective of this Course

- Introduce main concepts underlying machine learning with **mathematical rigor**.
  - No free lunch; Bias-variance trade-off; Stability; Generalization

- Uncover **mathematical principles** underlying various machine learning techniques.
  - Model selection; Regularization; Over-parametrization

- Show how to **theoretically** analyze learning algorithms.
  - Support vector machine; Deep neural networks

- Develop theory-oriented thinking which helps understand existing algorithms and create novel ones.

WARNING

This is a SERIOUS THEORY course

Taught in math – full of proofs and notations.
Try to make all ML folklores solid and rigorous.
Emphasis on theoretical foundations, not on techniques.

This is an ADVANCED ML course

Do not expect this to be a first course of ML.
Requires some math maturity and/or ML background.

“Should I take this course?”

- YES, if one of the following is true:
  - You have already taken a *serious* ML course
  - You have already taken a *serious* STAT course
  - You have some math maturity, enjoy developing theory, and have no problem with mathematical notations and proofs.
- NO, if one of the following is true:
  - You do not care about theory of ML and only want to know how machine learning algorithms work
  - You have little background in multi-variate calculus and probability
    - check: gradient, Hessian, Taylor expansion, conditional expectation, convergence
  - You do not want to spend at least 6 hours per week off the class
  - You do not want to work on very difficult homework that easily takes up to two days
- Welcome to talk to me if you cannot decide.

Logistics (1)

- **Lecturer** I-Hsiang Wang
  - email ihwang@ntu.edu.tw
  - office MD-524
  - office hours 17:30 – 18:30, Tues. and Wednesday

- **TA** Chen-Hao Hsiao
  - email r07942062@ntu.edu.tw
  - office hours 19:00 – 20:00, Thursday, BL-524

- **Time** 10:30 – 11:45, Tues. and Thursday
- **Location** EE2-106
- **Prerequisites** calculus, linear algebra, probability
- **Preferable** machine learning, optimization, analysis

Logistics (2)

- **Grading**
  - Homework (50%), Exam (25%), Project (25%)

- **Textbook**
  - N/A. Lectures will be based on my own slides and notes.
- **References**

[1] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: from Theory to Algorithms*, Cambridge University Press, 2014.

[2] Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, Kluwer Academic Publishers, 2004.

[3] Additional references: research papers and surveys to be assigned during lectures.

- **Website** NTUCOOL
  - Please turn on the notification and check your registered email regularly so that you do not miss anything important.

Topics – Statistical Principles (1/3)

- Unit 1: Introduction [1.5 weeks]
  - Probabilistic framework of machine learning [2/21]
  - Plug-in and universal consistency [2/21]
  - No-Free-Lunch Theorem [2/26]
  - Discriminative vs. generative approaches [2/26]
  - Learning linearly separable data via Perceptron [3/5]
  - Empirical risk minimization (ERM) [3/5]

Topics – Statistical Principles (2/3)

- Unit 2: Uniform convergence [3.5 weeks]
  - Probability toolkit: concentration inequalities [3/7]
  - Uniform law of large numbers [3/7]
  - Rademacher complexity [3/12]
  - Finite hypothesis class [3/12]
  - Bounds via growth function and VC dimension [3/14,3/19]
  - Bounds via covering number [3/21]
  - Margin-based bounds [3/26,3/28]

Topics – Statistical Principles (3/3)

- Unit 3: Stability and generalization [2 weeks]
  - The general learning framework [4/9]
  - Convex learning problems [4/9]
  - Stability and learnability [4/11]
  - Stability via regularization [4/16]
  - Information-theoretic notions of stability [4/18]

Topics – Algorithmic Principles (1/2)

- Unit 4: Algorithms [2 weeks]
  - Boosting [4/23]
  - Support vector machine and kernel methods [4/25]
  - Deep neural networks [4/30]
  - Validation and model selection [5/2]

Topics – Algorithmic Principles (2/2)

- Unit 5: Optimization [6 weeks]
    - The black-box model and oracle complexity [5/7]
    - Convex optimization for machine learning [5/9]
    - Convergence of gradient descent [5/14]
    - Accelerated gradient descent [5/16]
    - Mirror descent [5/21,5/23]
    - Stochastic gradient descent [5/28,5/30]
    - Stability of SGD [6/4]
    - Online to batch conversion [6/6]
    - Over-parametrization [6/11,6/13]

Homework Assignments

- In total 5 homework assignments (HW1–5). Each HW covers the materials in each unit. Problems released weekly.
  - A good strategy: work on the problems every week.
  - A bad strategy: work on the entire HW before the deadline …
- Late homework policy (X,Y: to be specified later)
  - due + X hours: × 0.5 due + Y hours: × 0.0
- Group work policy
  - You are allowed (and encouraged) to work in groups
  - Each group should be **≤ 3** people, and **groups are disjoint**
  - Put the student ID and the name of your partner(s) on the sheets
  - **Same partners** for the entire HW!
- **Plagiarism is not allowed.**
  - First time caught: that HW is graded 0.
  - Second time caught: semester grade F.

Notes, Slides, and Readings

- No textbook for this course. We use my own notes.
- The two main references are used for further readings.

- In class, I will use slides for better presentation
- Slides **do not cover all details**.
- Read the lecture notes for details such as proofs and further references.

- Readings are assigned on the website.
- Further readings are suggested in the notes.

- This course is still under development, so please tolerate delays in posting!

Project

- Theory-oriented projects
- Work in groups. # of people in each group: TBD
- Final presentation + report.
- Goal: overview some topics not covered in the lectures
  - Survey a certain topic in depth
  - Not just one paper; should be a series of papers.
  - Interpret with your own language and unify.
  - Exploration on unsolved open problems are welcome.
  - Experiments to validate the theoretical findings are welcome.
- Topics (suggested but not limited to):
  - Online learning; Reinforcement learning; Active learning; Unsupervised learning; Transfer learning; Deep learning; etc.

Tentative Schedule (1)

week	date	lecture	unit	note
1	02/19 02/21	Logistics and overview Probabilistic framework of machine learning	Introduction	
2	02/26 02/28	No-Free-Lunch Theorem Holiday (no lecture)	Introduction	
3	03/05 03/07	Learning linearly separable data via Perceptron Probability toolkit: concentration inequalities	Introduction	
4	03/12 03/14	Rademacher complexity	Uniform convergence	HW1 due
5	03/19 03/21	Growth function and VC dimension Covering number	Uniform convergence	
6	03/26 03/28	Margin-based bounds	Uniform convergence	
7	04/02 04/04	Spring break (no lecture)		HW2 due
8	04/09 04/11	General learning framework; Convex learning Stability and learnability	Stability and generalization	
9	04/16 04/18	Stability via regularization Information-theoretic notions of stability	Stability and generalization	

Tentative Schedule (2)

week	date	lecture	unit	note
10	04/23 04/25	Boosting Support vector machine and kernel methods	Algorithms	HW3 due
11	04/30 05/02	Deep neural networks Model selection and Validation	Algorithms	Project Proposal
12	05/07 05/09	The black-box model and oracle complexity Convex optimization for machine learning	Optimization	
13	05/14 05/16	Convergence of gradient descent Accelerated gradient descent	Optimization	HW4 due
14	05/21 05/23	Mirror descent	Optimization	
15	05/28 05/30	Stochastic gradient descent	Optimization	
16	06/04 06/06	Stability of SGD Online to batch conversion	Optimization	HW5 due
17	06/11 06/13	Over-parametrization	Optimization	
18	06/18 06/20	Exam Project Presentations	Finale	

Some Final Remarks

- WARNING (again):
  - This is a SERIOUS THEORY course.
  - This is an ADVANCED ML course.
  - Some “mathematical maturity” or background in ML is needed.
  - Loading is heavy.
  - This a graduate-level course. A high-quality project is anticipated.

- Auditing is welcome if capacity allows.

- Enroll in this class: there are still >10 spots left.
- The question is: do you really want to enroll.
- Sign up here:

  https://goo.gl/forms/Bf0lNU07CbvqfhZu2

Questions?